Portfolio Selection in Case of High Dimensionality

Master's Thesis submitted

to

Prof. Dr. Ostap Okhrin

Prof. Dr. Wolfgang Härdle

Humboldt-Universität zu Berlin

School of Business and Economics

Institute for Statistics and Econometrics

Ladislaus von Bortkiewicz Chair of Statistics

CASE - Center of Applied Statistics and Economics

by

Svetlana Bykovskaya

(562509)

in partial fulfillment of the requirements

for the degree of

Master of Economics and Management Science

Berlin, April 20, 2015

Abstract

The problem of portfolio selection has always been one of the most important topics in investment theory. This concerns not only the portfolio selection procedure itself but also the problems of mean and covariance matrix estimation. As nowadays large data sets are available, the problem of proper estimation in case of high dimensionality is becoming an important issue in financial analysis. The most serious difficulty is that most traditional estimation techniques have been developed for the case when the number of observations n is much larger than dimension of the data p and could not be applied in high-dimensional case (p is larger than n). In this thesis we analyze the efficiency of different portfolio selection methods and estimation techniques in case of high-dimensional data. We present an empirical study which compares performance of portfolios constructed of S&P 500 index components. The empirical analysis is performed in Matlab programming language.

Key words: portfolio selection, high dimensionality, shrinkage estimator, mean estimator, covariance matrix estimator

Zusammenfassung

Das Problem der Portfolioauswahl war immer eines der wichtigsten Themen in der Investionstheorie. Das betrifft nicht nur das Verfahren der Portfolioauswahl an sich aber auch die Probleme der Einschätzung des Erwartunswerts und der Kovarianz. Da grosse Datenmenge heutzutage verfügbar ist, wird das Problem der angemessenen Einschätzung im Falle der Hochdimensionalität ein wichtiges Thema in Finanzanalyse. Die allerwichtigste Schwierigkeit ist es, dass die am meisten traditionelle Einschtzungsmethoden für den Fall, wenn die Anzahl von Beobachtungen n viel grösser als Dimension von Daten p ist und auf hochdimensionalen Fall (p ist grösser als n) nicht angewendet werden kann, entwickelt wurden. In dieser Arbeit wird die Effizienz der unterschiedlichen Portfolioauswahlmethoden und Einschätzungsmethoden im Fall der hochdimensionalen Daten analysiert. Es wird eine empirische Untersuchung präsentiert, die Perfomanz der von S&P 500 Index Komponenten konstruierten Portfolios vergleicht. Die empirische Analyse ist in Matlab Programmsprache durchgeführt.

Schlüsselwörter: Portfolio Optimierung, Hochdimensionalität, Shrinkage Schätzer, Schätzer des Erwartunswerts, Schätzer der Kovarianz

Contents

List of Figures

List of Tables

1 Introduction

The problem of portfolio selection has always been one of the most important topics in investment theory. This concerns not only the portfolio selection procedure itself but also the problems of mean and covariance matrix estimation. As nowadays large data sets are available, the problem of proper estimation in case of high dimensionality is becoming an important issue in financial analysis. The most serious difficulty is that most traditional estimation techniques have been developed for the case when the number of observations n is much larger than dimension of the data p and could not be applied in high-dimensional case (p is larger than n).

In this paper we consider the following classical portfolio selection problems: global minimum variance problem, expected utility maximization problem and Sharpe ratio maximization problem. Global minimum variance problem considers an investor who wants to form the portfolio with the smallest variance. Two other problems imply making a compromise between getting higher income and lower risk. Global minimum variance method requires information on covariance matrix only, whereas expected utility maximization method and Sharpe ratio maximization problem method need also information on expected returns.

In order to find appropriate estimators of mean and covariance matrix, we analyze different estimation techniques which have recently been proposed in literature. Proper mean estimation in case of high dimensionality poses a significant challenge and there is known quite few appropriate mean estimators. The situation in case of covariance matrix estimation is significantly better as there are several approaches allowing to get well-performing estimators. The most popular estimation methods include factor models, shrinkage method and Bayesian and empirical Bayes estimators. Factor models assume that the data could be explained by some economic variables and have a factor structure. These models can significantly reduce the dimension of the data, which makes them very efficient when dealing with high-dimensional data. In this paper we concentrate on shrinkage technique which successfully competes with factor models.

We present an empirical study which analyzes the performance of portfolios constructed of S&P 500 index components and allows to compare different portfolio selection techniques and estimation methods. The empirical analysis is performed in Matlab programming language.

The paper is organized as follows. Section 2 describes theoretical framework of portfolio selection procedure. Sections 3 and 4 present high-dimensional mean and covariance matrix estimators. Section 5 provides an empirical study. Finally, Section 6 concludes.

2 Portfolio selection

The portfolio selection problem arises from the paper of **Markowitz (1952)**. Since then the topic has been extensively studied and developed into portfolio theory, one of the most important areas of financial analysis.

Portfolio selection procedure is based on two main problems: maximization of the expected wealth and minimization of the risk. Therefore, the investor faces the problem of forming the so-called mean-variance optimal portfolios. Significant contribution to the development of mean-variance analysis has been made by **Samuelson (1970), Jobson and Korkie (1980), Markowitz (1991)** and many other researchers (see also **Jobson and Korkie (1989), Golosnoy and Schmid (2007), Bodnar and Schmid (2009), Yu et al. (2009)**).

Markowitz defined the problem of portfolio selection as the minimization of portfolio risk for a given level of the expected return (the mean-variance optimization problem, see **Markowitz (1952)**). Since then there have been developed a number of portfolio selection methods which allow to construct a portfolio according to investor's preferences. We will concentrate on the following classical problems:

- Global minimum variance (GMV) problem,
- Expected utility maximization (EU) problem,
- Sharpe ratio maximization (Sharpe) problem.

In order to formalize these problems let us firstly define some theoretical framework. We consider an investor who wants to build a portfolio consisting of p assets. The main objective is to find optimal portfolio weights which meet investor's demands. Let $\mathbf{w}$ denote the $p-$dimensional vector of portfolio weights, the expected return of the portfolio and the covariance matrix are denoted by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. It is commonly assumed that $\boldsymbol{\Sigma}$ is positive definite. The portfolio optimization problems mentioned above could be described as maximization or minimization of some function depending on parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (subject to some additional constraints), in respect to $\mathbf{w}$. As in practice $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are typically unknown, we need to find appropriate estimators for them. The problems of estimation of the mean vector and the covariance matrix will be discussed in detail later.

Global minimum variance problem arises from the difficulty connected with the estimation of the mean vector of portfolio returns, which has been shown by **Merton (1980)**. In this case we concentrate on the minimization of the portfolio variance only. Therefore, we try to find the portfolio with the lowest return variance for a given covariance matrix $\mathbf{\Sigma}$. We can formalize the problem as follows:

$$\mathbf{w}'\mathbf{\Sigma}\mathbf{w} \to \min$$

$$\text{s.t.}$$

$$\mathbf{w}'\mathbf{1} = 1,$$

where $\mathbf{1}$ denotes the $p-$dimensional vector of ones.

The problem was considered in a great number of papers (see **Kempf and Memmel (2006), Okhrin and Schmid (2006), Bodnar and Schmid (2008); Bodnar, Parolya, Schmid (2014)**). By minimizing the portfolio variance, we obtain the following global minimum variance (GMV) portfolio weights:

$$\mathbf{w_{GMV}} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}}.$$

In case of expected utility maximization (EU) problem we want to maximize the mean-variance utility function of investor. This problem makes a compromise between maximization of the expected wealth and minimization of the risk, depending on investor's objectives. The EU problem can be formulated in the following way.

$$\mathbf{w}'\boldsymbol{\mu} - \frac{\alpha}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} \to \max$$

$$\text{s.t.}$$

$$\mathbf{w}'\mathbf{1} = 1,$$

where α is a risk-aversion coefficient of an investor. The problem has been widely discussed in literature (see **Ingersoll (1987), Okhrin and Schmid (2006); Bodnar, Parolya, Schmid (2013)**).

The weights of the expected utility (EU) optimal portfolio are given by

$$\mathbf{w_{EU}} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}} + \alpha^{-1}\boldsymbol{\mu}\mathbf{R},$$

where

$$\mathbf{R} = \mathbf{\Sigma}^{-1} - \frac{\mathbf{\Sigma}^{-1}\mathbf{1}\mathbf{1}'\mathbf{\Sigma}^{-1}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}}.$$

Sharpe ratio maximization problem deals with the Sharpe ratio, a quite popular measure of the portfolio performance (see **Cochrane (1999), Jobson and Korkie (1981), Memmel (2003), Ledoit and Wolf (2008)**). Sharpe ratio maximization problem is formulated as follows:

$$\frac{\mathbf{w}'\boldsymbol{\mu}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}} \to \max$$

s.t.

$$\mathbf{w}'\mathbf{1} = 1.$$

The Sharpe ratio (SR) optimal portfolio weights are given by

$$\mathbf{w_{SR}} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}.$$

3 High-dimensional mean estimators

With the development of portfolio theory the problem of expected return estimation has appeared. As it has already been mentioned, **Merton (1980)** described the difficulties arising from this problem and showed the importance of proper estimation of expected return. **James and Stein (1961)** showed the ineffectiveness of using sample mean estimator in case of multivariate normal distribution and suggested the new method of mean estimation.

There have been proposed a number of estimators which perform quite well (see, e.g., **Baranchik (1970), Berger and Bock (1976), Gleser (1986), Jorion (1986, 1991), Fourdrinier et al. (2003)**). However, most of them have been developed for the case when the number of observations n is much larger than the number of features p. As modern financial, biological and some other applications often require data analyzing in case of high dimensionality (p is larger than n), the problem of proper mean estimation is becoming increasingly important. We will concentrate on the following two techniques which have recently been proposed: James-Stein type estimator (**Chetelat and Wells (2012)**) and optimal linear shrinkage estimator (**Bodnar, Okhrin and Parolya (2015)**).

3.1 James-Stein type estimator

Let $\mathbf{Y}_n = \{\mathbf{Y}_{p1}, \dots, \mathbf{Y}_{pn}\}$ be the $p \times n$ sample of size n of a p-dimensional random vector distributed as $N_p(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, where the mean vector $\boldsymbol{\mu}_n$ and the covariance matrix $\boldsymbol{\Sigma}_n$ are unknown parameters. Let $\mathbf{S}_n$ be some estimator of $\boldsymbol{\Sigma}_n$ and

$$\mathbf{S}_n \sim Wishart_p(n, \boldsymbol{\Sigma}_n).$$

As appears from the definition of the Wishart distribution, in case of $p \leq n$ the estimator $\mathbf{S}_n$ is (almost surely) invertible, whereas in case of $p > n$ the estimator $\mathbf{S}_n$ is almost surely singular. Let also $\mathbf{I}$ denotes $p \times p$ identity matrix.

Our goal is to estimate $\boldsymbol{\mu}_n$ under the invariant quadratic loss

$$L(\boldsymbol{\mu}_n, \hat{\boldsymbol{\mu}}_n) = (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n)' \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_n).$$

The traditional estimator of $\boldsymbol{\mu}_n$ is a sample mean, which is given by

$$\overline{\mathbf{y}}_n = \frac{1}{n} \mathbf{Y}_n \mathbf{1}_n.$$

However, **Stein (1956)** showed that in case of $p \geq 3$ this estimator is inadmissible under quadratic loss $L(\boldsymbol{\mu}_n, \hat{\boldsymbol{\mu}}_n)$. He also proposed the idea of mean estimation, which has been improved later by James and some other researchers (see **James and Stein (1961), Efron and Morris (1972), Fourdrinier et al. (2003)**). As a result, there have been suggested the James—Stein type estimator

$$\hat{\boldsymbol{\mu}}_{n,JS} = (\mathbf{I} - \frac{(p-2)/(p-n+3)}{\overline{\mathbf{y}}_n' \mathbf{S}_n{}^{-1} \overline{\mathbf{y}}_n})\overline{\mathbf{y}}_n,$$

which dominates the sample mean estimator $\overline{\mathbf{y}}_n$. However, this estimator is appropriate only for the case $n \geq p$, because otherwise $\mathbf{S}_n$ is not invertible.

Chetelat and Wells (2012) extended this result, constructing a class of estimators for the case of $n < p$. We will consider one of these estimators only, a high-dimensional James—Stein type estimator

$$\hat{\boldsymbol{\mu}}_{n,JS(p>n)} = (\mathbf{I} - \frac{a\mathbf{S}_n\mathbf{S}_n^+}{\overline{\mathbf{y}}_n' \mathbf{S}_n^+ \overline{\mathbf{y}}_n})\overline{\mathbf{y}}_n,$$

where $\mathbf{S}_n^+$ is the Moore—Penrose generalized inverse of $\mathbf{S}_n$ (the Moore—Penrose generalized inverse is commonly used when the matrix is singular and, therefore, cannot be inverted.) This estimator dominates $\overline{\mathbf{y}}_n$ under invariant quadratic loss for all

$$0 \leq a \leq \frac{2(n-2)}{p-n+3}.$$

3.2 Optimal Shrinkage Estimator

In this section we will consider the optimal shrinkage estimator which has been proposed by **Bodnar, Okhrin and Parolya (2015).**

Let $\mathbf{Y}_n = \{\mathbf{Y}_{p1}, \ldots, \mathbf{Y}_{pn}\}$ be the $p \times n$ sample of size n of a p-dimensional random vector. We assume that

$$\mathbf{Y}_n \overset{d}{=} \boldsymbol{\Sigma}_n^{\frac{1}{2}} \mathbf{X}_n + \boldsymbol{\mu}_n \mathbf{1}_n',$$

where the mean vector $\boldsymbol{\mu}_n$ and the covariance matrix $\boldsymbol{\Sigma}_n$ are unknown parameters, $\mathbf{1}_n'$ is $n-$dimensional vector of ones and all $\mathbf{X}_n$ are independent and identically distributed with

$$E(\mathbf{X}_n) = 0, \quad Var(\mathbf{X}_n) = 1.$$

We use the following assumptions.

(A1) It exists $\lambda_0 > 0$ such that $\lambda_0 \leq \lambda_{min}(\boldsymbol{\Sigma}_n)$ uniformly on p, where $\lambda_{min}(\mathbf{A})$ denotes the smallest eigenvalue of the square matrix $\mathbf{A}$.

(A2) It exist $\gamma \in [0, 1), M_l > 0, M_u > 0$ such that the two limits $\lim_{p\to\infty} p^{-\gamma} \|\boldsymbol{\mu}_n\|^2 = M_n$ and $\lim_{p\to\infty} p^{-\gamma} \|\boldsymbol{\mu}_0\|^2 = M_0$ exist and it holds that $0 < M_l \leq M_0, M_n \leq M_u < \infty$.

The general linear shrinkage estimator of the mean vector is defined by

$$\hat{\boldsymbol{\mu}}_{GSE} = \alpha_n \overline{\mathbf{y}}_n + \beta_n \boldsymbol{\mu}_0,$$

where α_n and β_n are shrinkage intensities, and $\boldsymbol{\mu}_0$ is a shrinkage target satisfying **(A2)**. It should be noted, that the target vector $\boldsymbol{\mu}_0$ has to be independent of $\mathbf{Y}_n$ and, therefore, linearly independent of $\overline{\mathbf{y}}_n$.

In order to calculate the optimal shrinkage intensities for a given target vector $\boldsymbol{\mu}_0$, we have to minimize the following quadratic loss function

$$L(\boldsymbol{\mu}_n, \hat{\boldsymbol{\mu}}_{GSE}) = (\hat{\boldsymbol{\mu}}_{GSE} - \boldsymbol{\mu}_n)' \boldsymbol{\Sigma}_n^{-1} (\hat{\boldsymbol{\mu}}_{GSE} - \boldsymbol{\mu}_n).$$

Calculating the derivatives and the Hessian of L with respect to α_n and β_n, we obtain the following formulas for optimal shrinkage intensities:

$$\alpha_n^* = \frac{\overline{\mathbf{y}}_n' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n \boldsymbol{\mu}_0' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - \boldsymbol{\mu}_n' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 \overline{\mathbf{y}}_n' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0}{\overline{\mathbf{y}}_n' \boldsymbol{\Sigma}_n^{-1} \overline{\mathbf{y}}_n \boldsymbol{\mu}_0' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - (\overline{\mathbf{y}}_n' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0)^2},$$

$$\beta_n^* = \frac{\overline{\mathbf{y}}_n \boldsymbol{\Sigma}_n^{-1} \overline{\mathbf{y}}_n \boldsymbol{\mu}_n' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - \overline{\mathbf{y}}_n' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 \overline{\mathbf{y}}_n' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n}{\overline{\mathbf{y}}_n' \boldsymbol{\Sigma}_n^{-1} \overline{\mathbf{y}}_n \boldsymbol{\mu}_0' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0 - (\overline{\mathbf{y}}_n' \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_0)^2}.$$

As optimal shrinkage intensities α_n^* and β_n^* depend on unknown parameters $\boldsymbol{\Sigma}_n$ and $\boldsymbol{\mu}_n$, we have to construct appropriate estimators for them.

Let $\overline{\mathbf{y}}_n$ and $\mathbf{S}_n$ denote the sample mean vector and the sample covariance matrix:

$$\overline{\mathbf{y}}_n = \frac{1}{n} \mathbf{Y}_n \mathbf{1}_n,$$

$$\mathbf{S}_n = \frac{1}{n} (\mathbf{Y}_n - \overline{\mathbf{y}}_n \mathbf{1}_n')(\mathbf{Y}_n - \overline{\mathbf{y}}_n \mathbf{1}_n')' = \frac{1}{n} \mathbf{Y}_n \mathbf{Y}_n' - \overline{\mathbf{y}}_n \overline{\mathbf{y}}'_n.$$

Assume that **(A1)** and **(A2)** hold and $\frac{p}{n} \to c \in (0,1)$ for $n \to \infty$. Then the consistent estimators for α_n^* and β_n^* are given by

$$\hat{\alpha}_n^* = 1 - \frac{c}{1-c} \frac{\boldsymbol{\mu}_0' \mathbf{S}_n^{-1} \boldsymbol{\mu}_0}{(\frac{c^2}{1-c} + \overline{\mathbf{y}}_n' \mathbf{S}_n^{-1} \overline{\mathbf{y}}_n) \boldsymbol{\mu}_0' \mathbf{S}_n^{-1} \boldsymbol{\mu}_0 - (\boldsymbol{\mu}_0' \mathbf{S}_n^{-1} \boldsymbol{\mu}_0)^2}$$

and

$$\hat{\beta}_n^* = (1 - \hat{\alpha}_n^*) \frac{\overline{\mathbf{y}}_n' \mathbf{S}_n^{-1} \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0' \mathbf{S}_n^{-1} \boldsymbol{\mu}_0}.$$

Therefore, in case $c < 1$ we could define optimal linear shrinkage estimator for the mean vector by

$$\hat{\boldsymbol{\mu}}_{OLSE} = \hat{\alpha}_n^* \overline{\mathbf{y}}_n + \hat{\beta}_n^* \boldsymbol{\mu}_0.$$

This estimator provides almost surely smallest quadratic loss under large dimensional asymptotics.

As in case $c > 1$ the sample covariance matrix is not invertible, we need another estimator of $\mathbf{\Sigma}_n$. It seems to be reasonable to use the optimal linear shrinkage estimator suggested by **Bodnar, Gupta, Parolya (2014)** for the case $c \in (0, +\infty)$. This estimator is given by

$$\hat{\mathbf{\Sigma}}_{OLSE} = \hat{\alpha}^* \mathbf{\Sigma}_n + \hat{\beta}^* \mathbf{\Sigma}_0 \quad \text{with} \quad \|\mathbf{\Sigma}_0\|_{tr} \leq M,$$

where

$$\hat{\alpha}^* = 1 - \frac{\frac{1}{n}\|\boldsymbol{S}_n\|_{tr}^2 \|\mathbf{\Sigma}_0\|_F^2}{\|\boldsymbol{S}_n\|_F^2 \|\mathbf{\Sigma}_0\|_F^2 - (tr(\boldsymbol{S}_n \mathbf{\Sigma}_0))^2}$$

and

$$\hat{\beta}^* = \frac{tr(\boldsymbol{S}_n \mathbf{\Sigma}_0)}{\|\mathbf{\Sigma}_0\|_F^2}(1 - \hat{\alpha}^*).$$

The OLSE estimator (3.5) provides the asymptotic smallest Frobenius loss almost surely. As for the choice of $\mathbf{\Sigma}_0$, one could take, for instance, $\mathbf{\Sigma}_0 = \frac{1}{p}\mathbf{I}$.

Using this estimator for $\mathbf{\Sigma}_n$, we obtain the following bona fide estimator of the mean vector in case $c > 1$:

$$\hat{\boldsymbol{\mu}}_{OLSE}^1 = \hat{\alpha}_n^1 \overline{\mathbf{y}}_n + \hat{\beta}_n^1 \boldsymbol{\mu}_0,$$

where $\hat{\alpha}_n^1$ and $\hat{\beta}_n^1$ are given by

$$\hat{\alpha}_n^1 = 1 - \frac{c\boldsymbol{\mu}_0' \hat{\mathbf{\Sigma}}_{OLSE}^{-1} \boldsymbol{\mu}_0}{(c + \overline{\mathbf{y}}_n' \hat{\mathbf{\Sigma}}_{OLSE}^{-1} \overline{\mathbf{y}}_n)\boldsymbol{\mu}_0' \hat{\mathbf{\Sigma}}_{OLSE}^{-1} \boldsymbol{\mu}_0 - (\overline{\mathbf{y}}_n' \hat{\mathbf{\Sigma}}_{OLSE}^{-1} \boldsymbol{\mu}_0)^2}$$

and

$$\hat{\beta}_n^1 = (1 - \hat{\alpha}_n^1)\frac{\overline{\mathbf{y}}_n' \hat{\mathbf{\Sigma}}_{OLSE}^{-1} \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0' \hat{\mathbf{\Sigma}}_{OLSE}^{-1} \boldsymbol{\mu}_0}.$$

It should be noted, that although this method does not obviously provide the optimal shrinkage estimator, the estimator $\hat{\boldsymbol{\mu}}_{OLSE}^1$ dominates in most of cases the well-known estimators for the high-dimensional mean vector.

4 High-dimensional covariance matrix estimators

Another important problem appearing in portfolio selection procedure is the estimation of covariance matrix. As in the case of mean estimation, most covariance estimation methods have been developed in the framework of low dimensionality of the data (p is smaller than n). When using those estimators in case of high dimensionality, one could face the number of major challenges. These include, among others, the following problems:

- the covariance matrix is estimated with too much error,
- the estimated matrix is not invertible, and, therefore, one could not use traditional portfolio selection techniques.

Therefore, one needs to find appropriate covariance estimators which do not break down in case of high dimensionality. This problem has been widely discussed in literature (see **Haff (1980), Jagannathan and Ma (2003), Fan et al. (2008), Levina et al. (2008), Bai and Shi (2011), Bai and Li (2012)**). Basic methodologies which allow to construct covariance estimators taking account of a high-dimensional case could be classified into following categories.

- Factor models. These models assume that asset returns have a factor structure and, therefore, could be explained by some economic variables or factors. The main advantage of factor models is that they allow to reduce the high dimensionality of the data and, therefore, are efficient for analysis of large data sets. There is a lot of papers dealing with factors models (see **Fama and French (1993), Campbell, Lo, and MacKinlay (1997), Forni et al. (2000), Lettau and Ludvigson (2001), Bai and NG (2008)**).
- The shrinkage method. Shrinkage technique which has been proposed by **Stein (1956)** implies the estimation of the covariance matrix by an optimally weighted average of the sample estimator and another estimator. This method has been heavily discussed in literature (see **Muirhead (1987), Frost and Savarino (1986), Ledoit and Wolf (2012)**).
- Bayesian and empirical Bayes estimators. Unlike most estimation methods, Bayesian methods imply the use of not only data but also prior information concerning to covariance matrix (e.g. personal beliefs). More detailed information on these methods could

be found in **Chen (1979), Yang and Berger (1994), Sun and Berger (1998), Barnard et al. (2000)**.

We will concentrate on shrinkage estimation methods which have been proposed by **Ledoit and Wolf ((2003,a), (2003,b), 2004))**.

The covariance matrix shrinkage estimator is given by

$$\hat{\mathbf{\Sigma}} = \alpha\mathbf{F} + (1-\alpha)\mathbf{S},$$

where α is a shrinkage intensity, $\mathbf{F}$ is a shrinkage target, $\mathbf{S}$ is a sample covariance matrix.

Ledoit and Wolf suggested to choose as a shrinkage target some highly structured estimator. The main idea is a trade-off between bias and estimation error. The sample covariance matrix $\mathbf{S}$ cannot be taken as an estimator of $\mathbf{\Sigma}$ because it has a lot of estimation error. However, this estimator has a great advantage of being unbiased among others. On the other hand, when we consider estimator $\mathbf{F}$ with a lot of structure, it has a little estimation error, but could be significantly unbiased. Therefore, shrinkage $\mathbf{S}$ towards $\mathbf{F}$ allows to make a compromise between these two problems, providing at the same time acceptable estimation error and acceptable bias.

The problem of most shrinkage estimators is that their loss functions involve the inverse of the covariance matrix. Therefore, in case of high dimensionality these estimators break down. Ledoit and Wolf proposed a loss function which does not depend on this inverse.

Let us define the Frobenius norm of the $p \times p$ symmetric matrix $\mathbf{Z}$ with entries (z_{ij}), $i, j = 1, \ldots, p$:

$$\|\mathbf{Z}\|^2 = Trace(\mathbf{Z}^2) = \sum_{i=1}^{p}\sum_{j=1}^{p} z_{ij}^2.$$

The loss function is given by

$$L(\alpha) = \|\alpha\mathbf{F} + (1-\alpha)\mathbf{S} - \mathbf{\Sigma}\|^2.$$

In order to calculate the optimal shrinkage intensity we have to minimize the risk function $R(\alpha)$ with respect to α, which is calculated in the following way:

$$R(\alpha) = E(L(\alpha)) = \sum_{i=1}^{p}\sum_{j=1}^{p} E(\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij}))^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p} Var(\alpha f_{ij} + (1-\alpha)s_{ij}) + [E(\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij}))]^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha^2 Var(f_{ij}) + (1-\alpha)^2 Var(s_{ij}) + 2\alpha(1-\alpha)Cov(f_{ij}, s_{ij}) + (\alpha E(f_{ij} - s_{ij}) + E(s_{ij} - \sigma_{ij}))^2,$$

where

$$\mathbf{F} = (f_{ij}), \quad \mathbf{S} = (s_{ij}), \quad \mathbf{\Sigma} = (\sigma_{ij}).$$

Minimization of this loss function gives us the following formula for the optimal shrinkage intensity:

$$\alpha^* = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} Var(s_{ij}) - Cov(f_{ij}, s_{ij}) - E(f_{ij} - s_{ij})E(s_{ij} - \sigma_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{p} E[(f_{ij} - s_{ij})^2].}$$

The next challenge is to find appropriate shrinkage target **F**. We will concentrate on following three models which define **F**:

- Single-index model (**Ledoit and Wolf (2003,a)**),
- Constant correlation model (**Ledoit and Wolf (2003,b)**),
- Identity model (**Ledoit and Wolf (2004)**).

Matlab codes implementing these estimators are available for download on the website of Olivier Ledoit: **http://www.ledoit.net.**

4.1 Single-index model.

The single-index model which has been proposed by **Sharpe (1963)** assumes that stock returns are generated by:

$$x_{it} = \alpha_i + \beta_i x_{0t} + \epsilon_{it},$$

where x_{0t} are market returns. It is also assumed that:

$$Cov(\epsilon_{it}, x_{0t}) = 0,$$

$$Cov(\epsilon_{it}, \epsilon_{jt}) = 0, \quad i \neq j,$$

$$Var(\epsilon_{it}) = \delta_{ii}.$$

Therefore, the covariance matrix is defined by

$$\mathbf{\Phi} = \sigma_{00}^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{\Delta},$$

where σ_{00} is the variance of market returns, $\boldsymbol{\beta}$ is the vector of slopes and $\mathbf{\Delta}$ is the diagonal matrix containing residual variances δ_{ii}.

We use the following estimator for the covariance matrix of stock returns:

$$\mathbf{F} = s_{00}^2 \boldsymbol{b}\boldsymbol{b}' + \mathbf{D},$$

where s_{00} is the sample variance of market returns,

$$\boldsymbol{b} = (b_1, \dots b_p)'$$

and b_i is the least squares estimator for β_i, $\mathbf{D}$ is the diagonal matrix containing residual variance estimates d_{ii} which are based on the OLS residuals.

Ledoit and Wolf have been shown that this method allows to obtain invertible and well-conditioned estimator which is more efficient compared to many others estimators.

4.2 Constant correlation model.

Elton and Gruber (1973) suggested the constant correlation model which assumes that every pair of stocks has the same correlation coefficient.

Let y_{it} be the return on stock i during period t, where $i = 1, \dots, p$, $t = 1, \dots, n$. Therefore, we have to estimate $p+1$ parameters: the p individual variances and the constant correlation coefficient. We also assume that stock returns are independent and identically distributed over time and have finite fourth moments.

Firstly, we calculate the sample average of the returns of stock i in the following way:

$$\overline{y}_i = \frac{1}{n} \sum_{i=1}^{n} y_{it}.$$

Let $\mathbf{S} = (s_{ij})$ denote the sample covariance matrix. The sample correlations between the returns on stocks i and j are given by

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii} s_{jj}}}.$$

We calculate the average of sample correlations by the formula

$$\overline{r} = \frac{2}{(p-1)p} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} r_{ij}.$$

Our shrinkage target $\mathbf{F} = (f_{ij})$ is defined in the following way:

$$f_{ii} = s_{ii},$$

$$f_{ij} = \overline{r} \sqrt{s_{ii} s_{jj}}.$$

4.3 Identity model.

The identity model assumes shrinkage of sample covariance matrix towards an identity matrix:

$$\mathbf{F} = r\mathbf{I},$$

where $\mathbf{I}$ is an identity matrix and r is some coefficient. Therefore, in order to obtain optimal shrinkage covariance matrix estimator, one needs to find the optimal linear combination

$$\hat{\mathbf{\Sigma}} = \alpha(r\mathbf{I}) + (1 - \alpha)\mathbf{S}.$$

Ledoit and Wolf have been shown that this estimator has relatively small risk and is better-conditioned compared to sample covariance matrix.

5 Empirical study

In this section we compare different methods of portfolio selection using various mean and covariance estimation techniques.

5.1 Data description

We provide an empirical study based on American stock data. The portfolio is constructed from the components of S&P 500 index. The data on stock prices were extracted from the Yahoo Finance database which is provided on their website **http://finance.yahoo.com/.** All stock returns are denominated in US dollars.

We observe daily stock prices over the period from 27 July 2004 to 30 June 2014, which amounts to 2500 time observations. As for some stocks all data were unavailable, we excluded them and considered a portfolio consisting of 449 stocks. The list of them is provided in the appendix.

To provide some basic summary statistics for our data, we firstly construct an equally-weighted portfolio (EW). The weights of the EW portfolio are given by

$$\mathbf{w}_{EW} = (\frac{1}{p}, \dots, \frac{1}{p}),$$

where $\mathbf{w}_{EW}$ is a p-dimensional vector of portfolio weights, $p = 449$. We form an equally-weighted portfolio on 27 July 2004 and hold it unchanged until 30 June 2014.

Figure 1 shows the kernel density of daily equally-weighted portfolio prices for the whole period. Figure 2 illustrates the time-series of equally-weighted portfolio returns, which are expressed in percent per day. We can observe three sharp volatility spikes in the end of 2008, in the middle of 2010 and in the end of 2011, which reflect financial instability caused by financial crisis.

The descriptive statistics are presented as Table 1. Although portfolio returns vary within a large range of 24,89 %, the mean return of the portfolio is quite low and amounts to 0,027 %. As can be seen from Figure 2 and Table 1, the portfolio has a relatively high volatility.

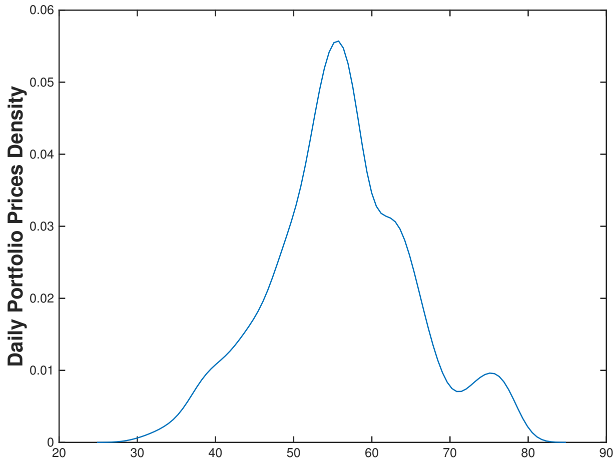

Figure 1: Kernel density of daily equally-weighted portfolio prices for the period 2004/07/27—2014/06/30.

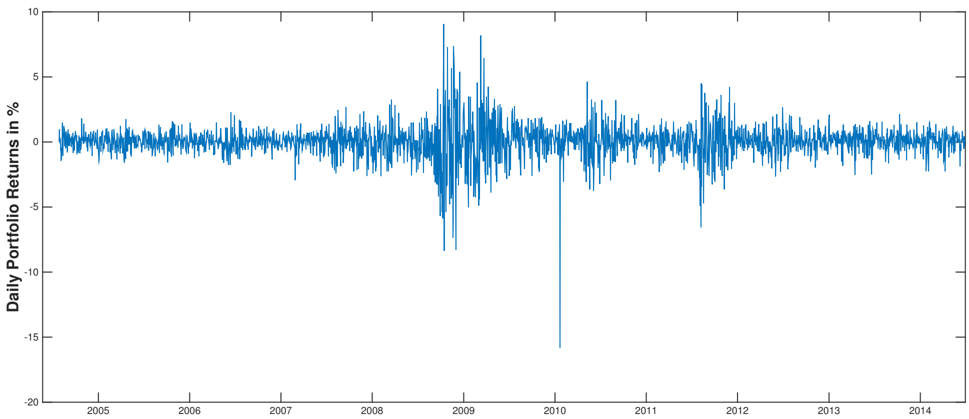

Figure 2: Daily equally-weighted portfolio returns, in %, for the period 2004/07/27—2014/06/30.

Variable	Mean	Min	Max	StD	Skewness	Curtosis
Daily Portfolio Returns	0.0272544	-15.82523	9.068881	1.261364	-0.9566974	19.90749

Table 1: Summary statistics of equally-weighted daily portfolio returns, in %, for the period 2004/07/27—2014/06/30.

5.2 Empirical set-up

In order to form a portfolio we use the following procedure:

- Firstly, we consider the data on stock returns over the in-sample period: from the day $(t-n)$ to the day t. We have chosen the size of the rolling window $n = 250$ that corresponds to 1 year. This estimation period allows us to take into account significant changes in the behavior of stock-market prices. Thus we have the case of high dimensionality:
$$c = \frac{p}{n} = 1,796.$$
- Using the data from the in-sample period, we estimate the mean vector $\boldsymbol{\mu}$ by the following estimators:
  - Sample mean estimator ($\hat{\boldsymbol{\mu}}_{sample}$),
  - James-Stein type estimator ($\hat{\boldsymbol{\mu}}_{JS}$),
  - Optimal shrinkage estimator ($\hat{\boldsymbol{\mu}}_{OLSE}$).
- Using the data from the in-sample period, we estimate the covariance matrix $\boldsymbol{\Sigma}$ by the following methods:
  - Single-index model ($\hat{\boldsymbol{\Sigma}}_m$),
  - Constant correlation model ($\hat{\boldsymbol{\Sigma}}_{cor}$),
  - Optimal shrinkage model ($\hat{\boldsymbol{\Sigma}}_{id}$).
- We compute optimal portfolio weights using the following portfolio selection techniques:
  - Global minimum variance (GMV) method,
  - Expected utility maximization (EU) method,
  - Sharpe ratio maximization (Sharpe) method.

- On the day $(t+1)$ we form the portfolio and hold it unchanged during the period of $k = 28$ days (out-of-sample period which corresponds to 1 month.) At the end of this period we rebalance the portfolio by repeating the whole procedure.

As a result, we have $T = 81$ samples:

- 80 samples over the period of 28 days,
- 1 sample over the period 10 days.

5.3 Comparison of mean estimators

Let us firstly compare different methods of mean estimation. As it has been already mentioned before, we use the following mean estimators: sample mean, James-Stein type estimator and optimal linear shrinkage estimator. As GMV portfolio doesn't require mean estimation, we will consider only EU and Sharpe portfolios. In order to represent the investor with medium risk aversion, we decided to construct EU portfolio with risk-aversion coefficient $\alpha = 0.00001$.

Figure 3 illustrates the time-series of EU portfolio returns for three mean estimators, where $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}_{cor}$. As can be seen from this figure, all three plots totally coincide. The reason could be that the returns are the same or differ only slightly. When comparing mean estimators for all the other portfolio selection methods and covariance estimators, one can observe the same. Therefore, in our case the choice of mean estimator does not make a great contribution to forming a portfolio. Nevertheless, we will try to choose "the best" mean estimator in every case.

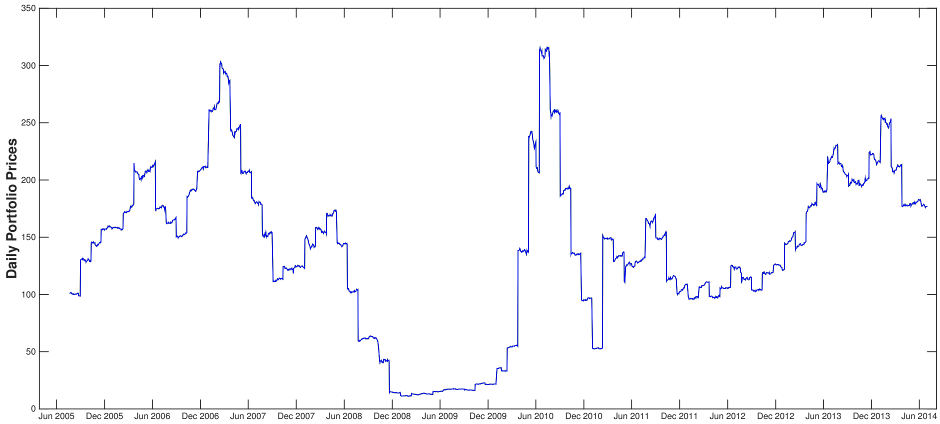

Figure 3: Daily EU portfolio prices for $\hat{\mathbf{\Sigma}}_{cor}$ and different mean estimators, for the period 2004/07/27—2014/06/30; red: sample mean $\hat{\boldsymbol{\mu}}_{sample}$, green: James-Stein type estimator $\hat{\boldsymbol{\mu}}_{JS}$, blue: optimal linear shrinkage estimator $\hat{\boldsymbol{\mu}}_{OLSE}$.

Let P_t be the price of the portfolio on the day t. We calculate the average monthly portfolio returns (AMR) in the following way.

- Monthly portfolio returns

$$MR_i = (\frac{P_{i*k+n+k}}{P_{i*k+n+1}} - 1), \tag{5.1}$$

  where we construct our portfolio on the day $(i * k + n + 1)$ and hold it until the day $(i * k + n + k)$.

- Average monthly portfolio returns

$$AMR = \frac{1}{T} \sum_{i=0}^{T-1} MR_i$$

Table 2 and Table 3 demonstrate AMR for EU and Sharpe portfolios. For every portfolio selection method and covariance matrix estimator we choose the mean estimator which provides higher AMR. One can see that for all portfolios it is an optimal linear shrinkage estimator $\hat{\boldsymbol{\mu}}_{OLSE}$. Later on we will take into account only this chosen mean estimator. As we use three covariance matrix estimators and three portfolio selection methods, we have to deal with 9 different portfolios.

Variable	EU, $\hat{\mathbf{\Sigma}}_{cor}$	EU, $\hat{\mathbf{\Sigma}}_{m}$	EU, $\hat{\mathbf{\Sigma}}_{id}$
$\hat{\boldsymbol{\mu}}_{sample}$	7,388421848	7,209776113	2,503959126
$\hat{\boldsymbol{\mu}}_{JS}$	7,388429400	7,209784118	2,503959634
$\hat{\boldsymbol{\mu}}_{OLSE}$	7,388797388	7,210216128	2,503965107

Table 2: Average monthly portfolio returns of EU portfolios, in %.

Variable	Sharpe, $\hat{\mathbf{\Sigma}}_{cor}$	Sharpe, $\hat{\mathbf{\Sigma}}_{m}$	Sharpe, $\hat{\mathbf{\Sigma}}_{id}$
$\hat{\boldsymbol{\mu}}_{sample}$	0,555045341	0,418689352	0,229641373
$\hat{\boldsymbol{\mu}}_{JS}$	0,555045473	0,418688478	0,229641321
$\hat{\boldsymbol{\mu}}_{OLSE}$	0,555517570	0,419116089	0,229655534

Table 3: Average monthly portfolio returns of Sharpe portfolios, in %.

5.4 Some basic portfolio characteristics

In order to analyze the structure of our portfolio we calculated some weight characteristics which can be found in Table 4.

It should be noted, that in case of $\hat{\mathbf{\Sigma}}_{id}$ for each portfolio both lowest and highest weights are relatively low and do not exceed 6 % (and 2 % in case of GMV and EU). One can see that in order to find a balance between high returns and low volatility Sharpe method requires significantly larger ranges of weights and a lot of short selling compared to EU and GMV methods.

The next step is to consider the performance of optimal shrinkage intensities which have been derived in order to estimate covariance matrices (see Section 4). Let us remind the formula of covariance matrix estimator:

$$\hat{\mathbf{\Sigma}} = \alpha\mathbf{F} + (1-\alpha)\mathbf{S},$$

where α is a shrinkage intensity, $\mathbf{F}$ is a highly structured estimator, $\mathbf{S}$ is a sample covariance matrix.

Figure 4 illustrates how those intensities change over the whole period. As one can see, for each covariance estimator shrinkage intensities are quite stable and vary from 0 to 0.09, which means that when constructing estimator the most part of weight is given to the sample covariance matrix S.

Variable	Lowest weight	Highest weight	Short interest
GMV, $\hat{\mathbf{\Sigma}}_{cor}$	-1,43074	15,73743	48,45268
GMV, $\hat{\mathbf{\Sigma}}_{m}$	-1,46130	14,98446	49,29291
GMV, $\hat{\mathbf{\Sigma}}_{id}$	-1,16346	1,06819	32,89080
EU, $\hat{\mathbf{\Sigma}}_{cor}$	-1,51045	15,50270	50,00011
EU, $\hat{\mathbf{\Sigma}}_{m}$	-1,56562	14,69653	50,53606
EU, $\hat{\mathbf{\Sigma}}_{id}$	-1,16495	1,07166	32,81353
Sharpe, $\hat{\mathbf{\Sigma}}_{cor}$	-16,14076	24,73560	398,78138
Sharpe, $\hat{\mathbf{\Sigma}}_{m}$	-18,07143	21,27886	400,78303
Sharpe, $\hat{\mathbf{\Sigma}}_{id}$	-3,80121	5,23037	158,21878

Table 4: Descriptive statistics for weights: averages of lowest and highest weights, averages of short interests, expressed in %. Short interest was calculated as the sum of securities sold short.

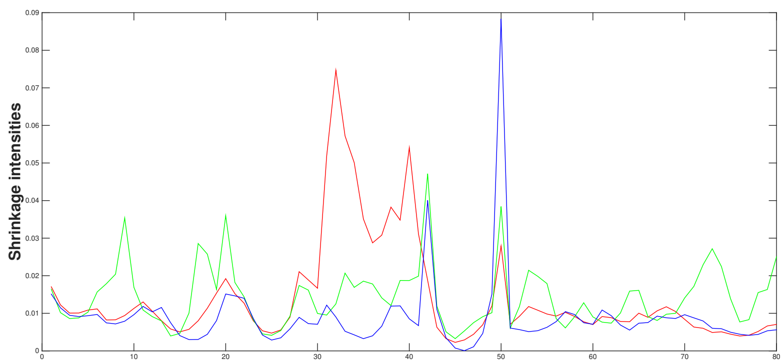

Figure 4: Shrinkage intensities for 3 covariance matrix estimators; red: $\hat{\mathbf{\Sigma}}_{cor}$, green: $\hat{\mathbf{\Sigma}}_{m}$, blue: $\hat{\mathbf{\Sigma}}_{id}$.

5.5 Portfolio returns

In order to estimate the effectiveness of different methods we will compare some basic portfolio characteristics. In this section we analyze the performance of our portfolios in terms of returns. For each portfolio we calculated monthly portfolio returns MR_i, using the formula (5.1). The plots and descriptive statistics are presented as Figure 5 and Table 5.

One can observe that for all three types of portfolio the estimator $\hat{\mathbf{\Sigma}}_{cor}$ provides returns which are slightly higher than the returns provided by the estimator $\hat{\mathbf{\Sigma}}_m$. At the same time, the returns in case of $\hat{\mathbf{\Sigma}}_{cor}$ vary within not so large range as in case of $\hat{\mathbf{\Sigma}}_m$, providing more stable returns (except for the case of Sharpe portfolio, where the range is significantly larger).

We consider now returns which are expressed as percentage, therefore their standard deviations are also expressed as percentages, reflecting the changes in portfolio returns. It should be noted that the GMV procedure minimizes not those variances, but variances reflecting the changes in portfolio prices. This explains the fact, that GMV portfolios have the highest volatilities compared to EU and Sharpe portfolios, at the same time providing the highest returns. Sharpe method of portfolio selection demonstrate the opposite trend, constructing portfolios with lowest returns and the lowest volatility. As for EU portfolios, they make a compromise between the maximization of returns and minimization of volatility, providing medium returns and medium volatility.

For each type of portfolio the situation in case of the estimator $\hat{\mathbf{\Sigma}}_{id}$ considerably differs from the situations in case of $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_m$. The returns provided by $\hat{\mathbf{\Sigma}}_{id}$ method are significantly lower than in cases of $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_m$. Therefore, when comparing covariance methods, one can see that portfolios in case of $\hat{\mathbf{\Sigma}}_{id}$ estimator are less volatile, providing not very high but stable income compared to other covariance estimators. This could be explained by the fact, that $\hat{\mathbf{\Sigma}}_{id}$ method assumes shrinkage to an identity matrix, which does not reflect the correlation between the stock returns and, therefore, $\hat{\mathbf{\Sigma}}_{id}$ is more structured but less informative. At the same time, both $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_m$ methods use the models which take into account those correlation and, therefore, reflect the real situation in a more adequate way.

Summing up, we can say that there is no big difference between $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_m$ methods as they provide quite similar monthly returns. However, $\hat{\mathbf{\Sigma}}_{cor}$ performs slightly better: it provides higher returns and at the same time lower standard deviations. As for portfolio selection techniques, one can observe that a risk-averse investor would rather prefer Sharpe method to GMV and EU methods. Obviously, among all three portfolio types, GMV portfolio allows to get the highest income, athough it implies a lot of risk.

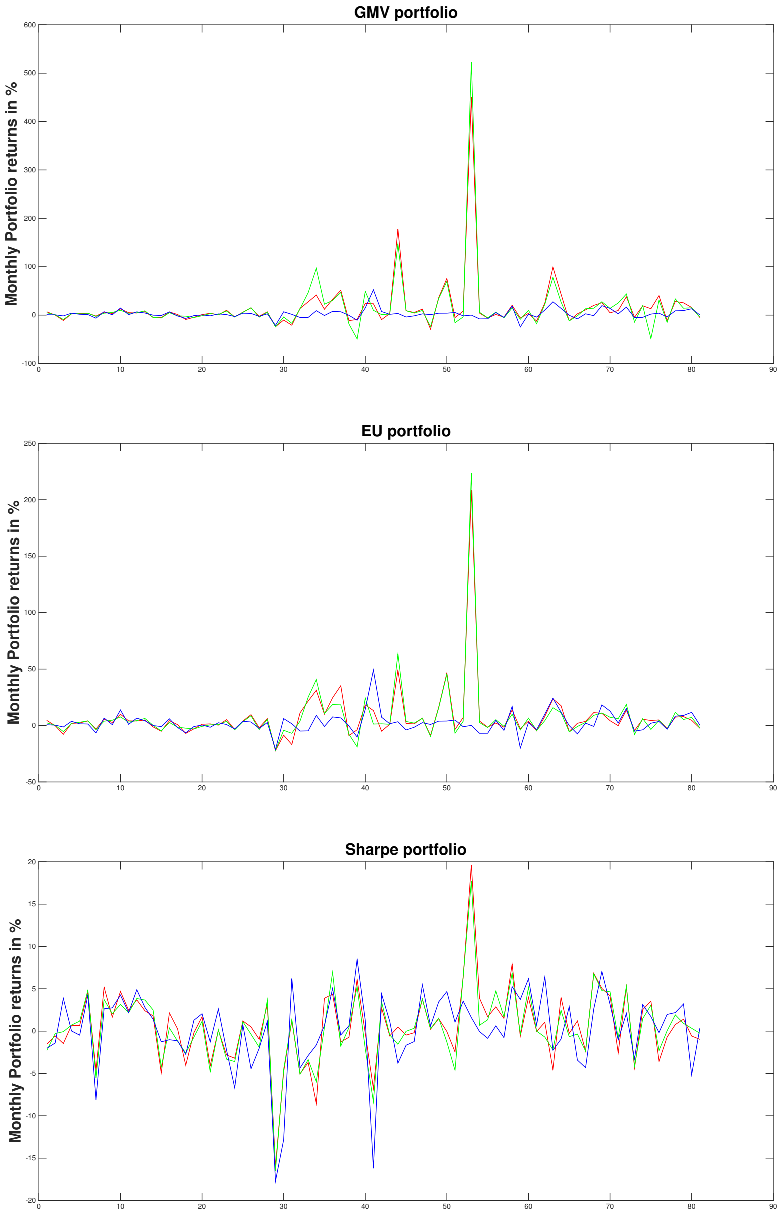

Figure 5: Montly portfolio returns for GMV, EU and Sharpe portfolio, for different covariance matrix estimators, expressed in %; red: $\hat{\mathbf{\Sigma}}_{cor}$, green: $\hat{\mathbf{\Sigma}}_{m}$, blue: $\hat{\mathbf{\Sigma}}_{id}$.

Variable	Mean	Min	Max	StD
GMV, $\hat{\mathbf{\Sigma}}_{cor}$	16,49495	-28,89790	450,03243	55,97677
GMV, $\hat{\mathbf{\Sigma}}_{m}$	15,99857	-49,18820	522,47827	63,35308
GMV, $\hat{\mathbf{\Sigma}}_{id}$	2,63649	-24,45987	52,20504	9,66348
EU, $\hat{\mathbf{\Sigma}}_{cor}$	7,38880	-22,25286	208,44697	25,39034
EU, $\hat{\mathbf{\Sigma}}_{m}$	7,21022	-21,85493	223,83376	27,20377
EU, $\hat{\mathbf{\Sigma}}_{id}$	2,50397	-21,11880	49,06197	8,92902
Sharpe, $\hat{\mathbf{\Sigma}}_{cor}$	0,55552	-15,97724	19,64814	4,32982
Sharpe, $\hat{\mathbf{\Sigma}}_{m}$	0,41912	-16,47523	17,73390	4,20370
Sharpe, $\hat{\mathbf{\Sigma}}_{id}$	0,22966	-17,70625	8,47515	4,48553

Table 5: Descriptive statistics for monthly portfolio returns, expressed in %.

5.6 Portfolio volatility

In this section we will compare the performance of our portfolios in terms of volatility. First of all let us concentrate on portfolio standard deviations.

For each portfolio for each of $T = 81$ samples we computed the standard deviation. The results are presented as Figure 6. Additionally, Table 6 provides descriptive statistics for these standard deviations. As one can see, for each portfolio type the plots for covariance matrix estimators $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_{m}$ (red and green lines, respectively) are quite similar.

Let us consider the cases of GMV and EU portfolios. We can observe large spikes in case of $\hat{\mathbf{\Sigma}}_{cor}$ as well as in case of $\hat{\mathbf{\Sigma}}_{m}$. Although their descriptive statistics are very similar, the estimator $\hat{\mathbf{\Sigma}}_{m}$ provides slightly higher volatility than the estimator $\hat{\mathbf{\Sigma}}_{cor}$. As for $\hat{\mathbf{\Sigma}}_{id}$, portfolios in this case are less volatile and have no significant bursts of volatility.

In case of Sharpe portfolio another picture is observed. We can observe significantly lower volatilities compared to GMV and EU portfolios. Furthermore, we can see that $\hat{\mathbf{\Sigma}}_{id}$ provides some large spikes comparable to spikes in case of $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_{m}$.

Generally speaking, GMV portfolios provide the highest volatility, EU portfolios provide the middle volatility and Sharpe portfolios provide the lowest volatility (except for the case GMV, $\hat{\mathbf{\Sigma}}_{id}$). For every type of portfolio the estimator $\hat{\mathbf{\Sigma}}_{id}$ gives us portfolio with minimum variance.

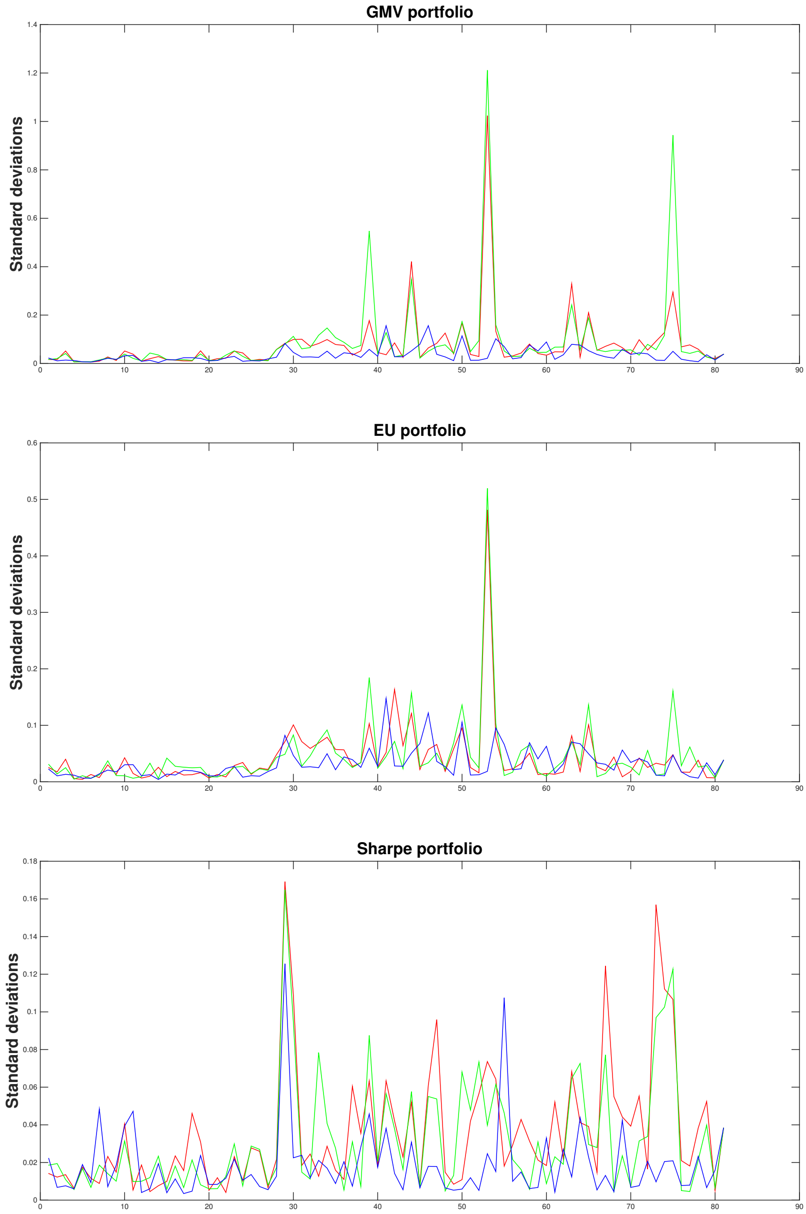

Figure 6: Standard deviations for GMV, EU and Sharpe portfolio, for different covariance matrix estimators, expressed in %; red: $\hat{\mathbf{\Sigma}}_{cor}$, green: $\hat{\mathbf{\Sigma}}_{m}$, blue: $\hat{\mathbf{\Sigma}}_{id}$.

Variable	Mean	Min	Max	StD
GMV, $\hat{\mathbf{\Sigma}}_{cor}$	0,07580	0,00618	1,02365	0,12742
GMV, $\hat{\mathbf{\Sigma}}_{m}$	0,08943	0,00660	1,21148	0,17713
GMV, $\hat{\mathbf{\Sigma}}_{id}$	0,03520	0,00399	0,15647	0,03042
EU, $\hat{\mathbf{\Sigma}}_{cor}$	0,04200	0,00442	0,48156	0,05803
EU, $\hat{\mathbf{\Sigma}}_{m}$	0,04489	0,00472	0,51978	0,06483
EU, $\hat{\mathbf{\Sigma}}_{id}$	0,03250	0,00424	0,14760	0,02690
Sharpe, $\hat{\mathbf{\Sigma}}_{cor}$	0,03613	0,00409	0,16920	0,03362
Sharpe, $\hat{\mathbf{\Sigma}}_{m}$	0,03173	0,00459	0,16480	0,03076
Sharpe, $\hat{\mathbf{\Sigma}}_{id}$	0,01857	0,00350	0,12556	0,01948

Table 6: Descriptive statistics for standard deviations, expressed in %.

Secondly, we will consider relative standard deviations (RSD) which have been calculated in the following way.

Let $\mathbf{v}_i$ be the vector of portfolio returns over out-of-sample period, $i = 1, \ldots, T$. Then

$$RSD_i = \frac{std(\mathbf{v}_i)}{mean(\mathbf{v}_i)}. \tag{5.2}$$

For each portfolio for each of $T = 81$ samples we computed the relative standard deviation. Figure 7 and Table 7 present the time-series of RSD and the descriptive statistics for all 9 portfolios.

One can see that among all the portfolios GMV portfolios have the largest relative standard deviations, which could vary within significantly large ranges: 18 % in case of $\hat{\mathbf{\Sigma}}_{id}$, and more than 50 % in cases of $\hat{\mathbf{\Sigma}}_{m}$ and $\hat{\mathbf{\Sigma}}_{id}$. As for EU portfolios, ranges are not so large and do not exceed 26 %. Sharpe portfolios have the narrowest range of 6-9%.

We can observe that for each of $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_{m}$ covariance estimators the plots for EU and GMV portfolios are very similar. Namely, one can see that RSD means and other descriptive statistics of GMV portfolios are equal to doubled ones of EU portolios. As for $\hat{\mathbf{\Sigma}}_{id}$ estimator, the characteristics for EU and GMV portfolios are quite similar.

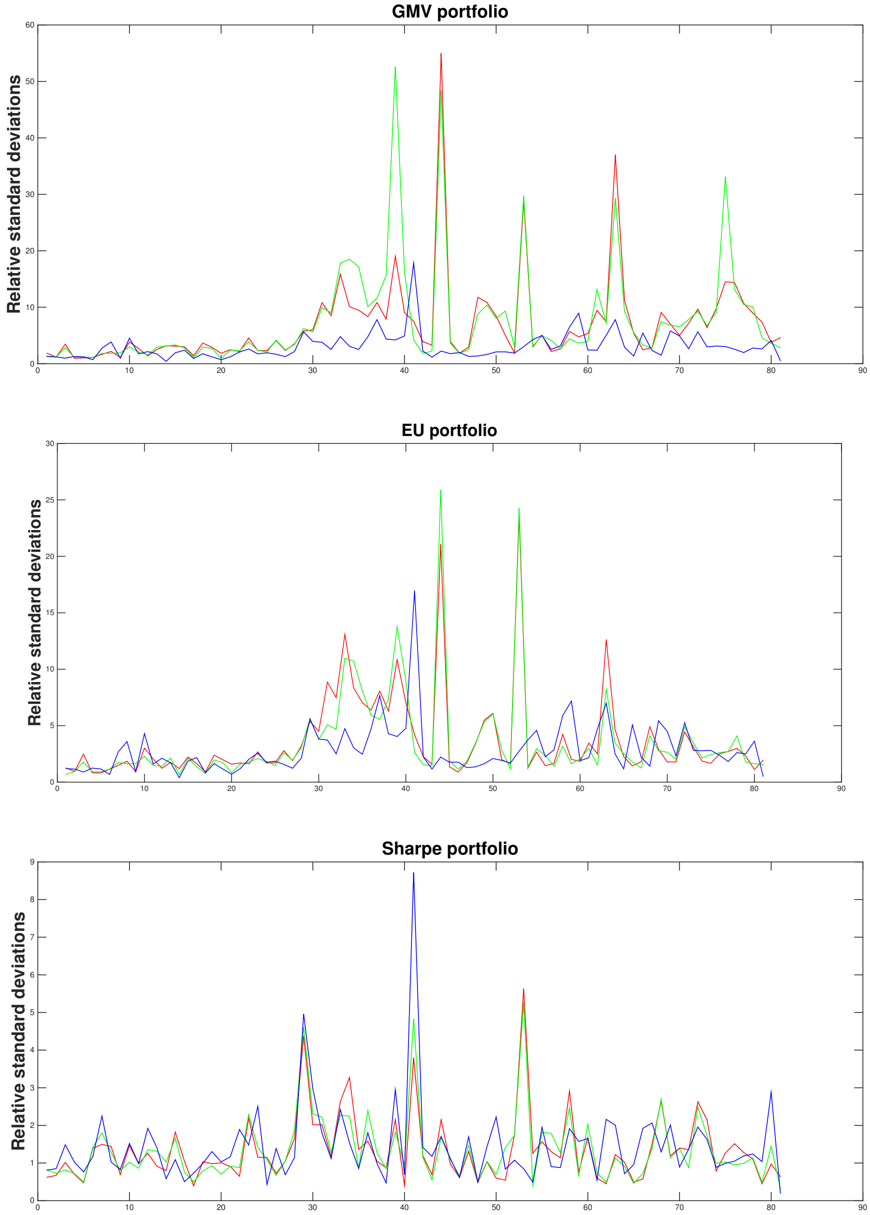

Figure 7: Relative standard deviations for GMV, EU and Sharpe portfolio, for different covariance matrix estimators, expressed in %.; red: $\hat{\mathbf{\Sigma}}_{cor}$, green: $\hat{\mathbf{\Sigma}}_{m}$, blue: $\hat{\mathbf{\Sigma}}_{id}$.

Variable	Mean	Min	Max	StD
GMV, $\hat{\mathbf{\Sigma}}_{cor}$	6,88200	0,94733	54,9829	7,92642
GMV, $\hat{\mathbf{\Sigma}}_{m}$	7,61939	0,93243	52,6204	9,40473
GMV, $\hat{\mathbf{\Sigma}}_{id}$	3,03113	0,40993	17,8137	2,42305
EU, $\hat{\mathbf{\Sigma}}_{cor}$	3,67682	0,81352	23,63765	3,97938
EU, $\hat{\mathbf{\Sigma}}_{m}$	3,58473	0,67882	25,91581	4,25597
EU, $\hat{\mathbf{\Sigma}}_{id}$	2,84916	0,39890	16,94780	2,25907
Sharpe, $\hat{\mathbf{\Sigma}}_{cor}$	1,34858	0,39300	5,62933	0,89709
Sharpe, $\hat{\mathbf{\Sigma}}_{m}$	1,34781	0,31486	5,27900	0,90422
Sharpe, $\hat{\mathbf{\Sigma}}_{id}$	1,44762	0,19805	8,72152	1,09618

Table 7: Descriptive statistics for relative standard deviations, expressed in %.

As it can be seen from Table 7, in cases of EU and GMV portfolios $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_{m}$ estimators provide higher RSD means compared to $\hat{\mathbf{\Sigma}}_{id}$ estimator. In case of Sharpe portfolio one can observe the opposite picture.

In conclusion, in order to get the lowest standard deviations as well as the lowest relative standard deviations, one should prefer Sharpe type of portfolio to GMV and EU portfolios, and the covariance matrix estimator $\hat{\mathbf{\Sigma}}_{id}$ to estimators $\hat{\mathbf{\Sigma}}_{cor}$ and $\hat{\mathbf{\Sigma}}_{m}$.

5.7 High dimensionality versus low dimensionality

Up to that point we analyzed the performance of portfolios which have been formed in case of high dimensionality. In order to compare the effectiveness of the different estimators and portfolio selection techniques we will consider now the case of low dimensionality. For this we repeat the whole procedure which has been described in section 5.2 given to $n = 500$, which corresponds to two years. Therefore, we have the case of low dimensionality:

$$c = \frac{p}{n} = 0,898.$$

When comparing portfolio returns for different mean estimators, one can observe that the returns are almost the same as in case of high dimensionality. As earlier, for each portfolio the optimal linear shrinkage estimator performs slightly better than the other mean estimators. Therefore, we will consider only those 9 portfolios which have been constructed using this estimator.

In order to compare cases of high and low dimensionality we will compute some basic portfolio characteristics and present them as a table.

The first descriptive statistic AAR which represents portfolio returns could be calculated in the following way.

- Daily returns
$$R_i = \frac{P_{i+1}}{P_i} - 1$$

- Monthly averaged returns
$$r_i = \frac{1}{k}\sum_{j=1}^{k} R_{i*k+n+j}$$

- Annualized averaged returns
$$ar_i = \frac{1}{12}\sum_{j=1}^{12}(1 + r_{i*12+j})^{12} - 1$$

- Average of annualized averaged returns
$$AAR = \frac{1}{m}\sum_{i=1}^{m} ar_i,$$
where m is a number of years.

The next two descriptive statistics will represent the portfolio volatility. Let $\mathbf{v}_i$ be the vector of portfolio returns over out-of-sample period, $i = 1, \ldots, T$.

- Average of annualized standard deviations
$$StD = \frac{1}{T}\sum_{i=0}^{T-1} std(\mathbf{v}_i) * \sqrt{12},$$
where $std(\mathbf{v}_i)$ is a standard deviation of returns during the out-of-sample period i.

- Average of relative standard deviations
$$RStD = \frac{1}{T}\sum_{i=0}^{T-1} RSD_i,$$
where RSD_i is a relative standard deviation of returns during the out-of-sample period i (see formula (6.2)).

The descriptive statistics mentioned above are presented as Table 8.

Let us firstly concentrate on the case of high dimensionality. Now we compare portfolios in terms of daily portfolio returns (AAR), and the situation differs from those we observed

	$n = 250, c = 1.796$			$n = 500, c = 0.898$		
Variable	AAR	StD	RStD	AAR	StD	RStD
GMV, $\hat{\mathbf{\Sigma}}_{cor}$	10,52112	26,25829	6,88200	53,21163	61,95880	13,27971
GMV, $\hat{\mathbf{\Sigma}}_{m}$	21,41505	30,98001	7,61939	20,95107	38,51386	12,01166
GMV, $\hat{\mathbf{\Sigma}}_{id}$	1,75008	12,19263	3,03113	1,02721	10,71981	3,55060
EU, $\hat{\boldsymbol{\mu}}_{OLSE}$, $\hat{\mathbf{\Sigma}}_{cor}$	2,52315	14,55033	3,67682	3,00172	16,88183	6,40448
EU, $\hat{\boldsymbol{\mu}}_{OLSE}$, $\hat{\mathbf{\Sigma}}_{m}$	3,08695	15,55061	3,58473	2,83048	16,23789	6,06982
EU, $\hat{\boldsymbol{\mu}}_{OLSE}$, $\hat{\mathbf{\Sigma}}_{id}$	1,44515	11,25853	2,84916	0,87620	10,11949	3,39006
SR, $\hat{\boldsymbol{\mu}}_{OLSE}$, $\hat{\mathbf{\Sigma}}_{cor}$	2,34392	12,51622	1,34858	2,77375	11,45020	1,38876
SR, $\hat{\boldsymbol{\mu}}_{OLSE}$, $\hat{\mathbf{\Sigma}}_{m}$	1,92749	10,99104	1,34781	2,91033	12,00709	1,37930
SR, $\hat{\boldsymbol{\mu}}_{OLSE}$, $\hat{\mathbf{\Sigma}}_{id}$	0,53161	6,43121	1,44762	0,58121	4,54455	1,46853

Table 8: Average annualized returns, annualized standard deviations and averaged relative standard deviations for cases $c > 1$ and $c < 1$, expressed in %, 2004-2014

in section 5.5 (where we considered monthly portfolio returns). Namely, here $\hat{\mathbf{\Sigma}}_m$ estimation method seems to be more profitable compared to $\hat{\mathbf{\Sigma}}_{cor}$ estimation method.

Now we will concentrate on comparison of the cases of high and low dimensionality. First of all, one can observe that the returns (AAR) in both cases are quite similar for all portfolios (except for GMV, $\hat{\mathbf{\Sigma}}_{cor}$).

Let us consider GMV and EU portfolios. Interestingly, in case of high dimensionality $\hat{\mathbf{\Sigma}}_m$ provides higher returns than $\hat{\mathbf{\Sigma}}_{cor}$, whereas in case of low dimensionality the situation is opposite. As for Sharpe portfolios, for $c > 1$ the estimator $\hat{\mathbf{\Sigma}}_m$ provides lower returns than $\hat{\mathbf{\Sigma}}_{cor}$, and the opposite is true for the case of $c < 1$. It should be noted, that for all portfolios the trend for $\hat{\mathbf{\Sigma}}_{id}$ is the same in both cases: it provides significantly lower returns and lower volatility compared to other estimators.

We can also observe that both StD and RStD are considerably higher in case of low dimensionality for GMV and EU portfolios, whereas for Sharpe portfolios they are almost the same.

Summing up, we can say that when comparing the performance of portfolios in cases of high and low dimensionality, we can describe all the trends in quite a similar way if we reverse the order of $\hat{\mathbf{\Sigma}}_m$ and $\hat{\mathbf{\Sigma}}_{cor}$ in case of low dimensionality.

6 Conclusion

The aim of this thesis is to analyze the portfolio selection problem in case of high-dimensional data (when the number of observations n is much larger than the number of features p). This problem addresses the following issues: the estimation of mean and covariance matrix and portfolio selection procedure. As most traditional estimators have been developed in framework of low dimensionality and break down in case of high-dimensional data, the problem of estimation of high-dimensional mean and high-dimensional covariance matrix is becoming increasingly important topic in financial analysis.

Firstly, we concentrated on portfolio allocation procedure. There have been considered the following portfolio selection problems: global minimum variance problem, expected utility maximization problem and Sharpe ratio maximization problem. The first one requires the estimation of covariance matrix only, while other two methods imply the estimation of both mean and covariance matrix.

As for high-dimensional mean estimators, we considered the following two techniques which have recently been proposed: James-Stein type estimator and optimal linear shrinkage estimator. As regards high-dimensional covariance matrix estimators, we concentrated on shrinkage method which implies the estimation of covariance matrix by an optimally weighted average of sample estimator and another highly structured estimator. This technique allows to cope with two main problems of sample covariance estimator: it has too much error and cannot be inverted. Ledoit and Wolf suggested three different models for constructing highly structured estimator: constant correlation model, single-index model and identity model. The first two estimators contain the information on portfolio variance, whereas the last one is an identity matrix multiplied by some coefficient.

In order to test the efficiencies of estimation techniques and portfolio selection methods we presented an empirical study in which we consider portfolios constructed from S&P 500 index components. Surprisingly, when comparing portfolios which have been formed using sample mean, James-Stein type estimator and optimal linear shrinkage estimator, one can observe that the returns are almost the same and, therefore, in this case there is no big difference between those estimators. However, it should be noted that the optimal linear shrinkage estimator provides slightly higher returns compared to other estimators.

When analyzing the performance of different covariance matrix estimators, one can see that the constant correlation model provides higher returns and lower standard deviations compared to single-index model, although both techniques form quite similar portfolios. However, both estimation methods imply a lot of risk. In contrast, the returns in case of identity model are significantly lower, whereas the risk is very small. This difference is explained by different frameworks of the models: first two estimators contain more information on portfolio variance and have not so much of structure, while the last one is more structured, but reflects the real situation less accurate.

The analysis of different portfolio allocation methods showed the following: Sharpe portfolio are more preferable for risk-averse investor, whereas expected utility maximization portfolio and global minimum variance portfolio provide significantly higher income but are much more risky.

To sum up, we can see that different combinations of portfolio selection and estimation techniques allow to form a variety of portfolios which are suitable for different types of investors. Futhermore, if an investor wants to increase or decrease the risk, there are available different ways to do it: he could either change the portfolio allocation method or choose another method of covariance matrix estimation.

Further research in this area implies, for example, considering other mean and covariance matrix estimation techniques and use of more sophisticated portfolio selection methods.

7 Bibliography

Bai, J. and Li, K.P. (2012), Statistical analysis of factor models of high dimension. *The Annals of Statistics, 40(1):* 436—465

Bai, J, and Ng, S. (2008) Large Dimensional Factor Models. *Foundations and Trends in Econometrics, 3(2):* 89—163.

Bai, J., Shi, S. (2011) Estimating high dimensional covariance matrices and its applications. *Department of Economics Columbia University, Discussion Paper No.: 1112—03*

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *The Annals of Mathematical Statistics, 41(2):* 642—645.

Barnard, J. R. McCulloch, and Meng, X.L. (2000) Covariance Matrices In Terms Of Standard Deviations And Correlations, With Application To Shrinkage. *Statistica Sinica 10:* 1281—1311.

Berger, J. O. and Bock, M. E. (1976). Combining independent normal mean estimation problems with unknown variances. *The Annals of Statistics, 4(3):* 642—648.

Bodnar, T., Schmid, W. (2008). Estimation of optimal portfolio compositions for gaussian returns. *Statistics & Decisions, 26:* 179—201.

Bodnar, T., Schmid, W. (2009). Econometrical analysis of the sample efficient frontier. *The European Journal of Finance 15:* 317—335.

Bodnar, T., Parolya, N., Schmid, W. (2013). On the equivalence of quadratic optimization problems commonly used in portfolio theory. *European Journal of Operational Research, 229:* 637—644.

Bodnar, T., Parolya, N., Schmid, W. (2014). Estimation of the Global Minimum Variance Portfolio in High Dimensions. *ArXiv preprint 1406.0437v1.*

Bodnar, T., Gupta, A., Parolya, N. (2014). On the Strong Convergence of the Optimal Linear Shrinkage Estimator for Large Dimensional Covariance Matrix. *ArXiv preprint 1308.2608v2.*

Bodnar, T., Okhrin, O., Parolya, N. (2015). Optimal Shrinkage Estimator for High Dimensional Mean Vector. *Forthcoming.*

Campbell, J.Y., A. Lo, and MacKinlay, A.C. (1997) The Econometrics of Financial Markets *Princeton University Press, New Jersey.*

Chen, C. F. (1979). Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *Journal of the Royal Statistical Society, Series B 41:* 235—248.

Chetelat, D. and Wells, M. (2012). Improved multivariate normal mean estimation with unknown covariance when p is greater than n. *The Annals of Statistics, 40(6):* 3137—3160.

Cochrane, J.H. (1999.) Portfolio advice for a multifactor world. *NBER working paper no. w7170.*

Efron, B. and Morris, C. (1973) Steins estimation rule and its competitorsan empirical bayes approach. *Journal of the American Statistical Associa- tion, 68(341)*:117—130.

Elton, E.J., Gruber, M.J. (1973) Estimating the dependence structure of share prices. *Journal of Finance, 28:* 1203—1232.

Ingersoll, J.E. (1987). Theory of Financial Decision Making. *Rowman & Littlefield Publishers.*

Fama, E. F., French, K. R. (1993) Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics, 33:* 3C56.

Fan, J., Fan Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147.1:* 186—197.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000) The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics,82:* 540—554.

Fourdrinier, D., Strawderman, W. E. and Wells, M. T. (2003). Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix. *Journal of Multivariate Analysis, 85* 24—39.

Frost, P.A., Savarino, J.E., (1986) An empirical Bayes approach to portfolio selection. *Journal of Financial and Quantitative Analysis, 21:* 293—305.

Gleser, L. J. (1986). Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *The Annals of Statistics, 14(4):* 1625—1633.

Golosnoy, V., Schmid, W. (2007) EWMA control charts for monitoring optimal portfolio weights. *Sequential Analysis: Design Methods and Applications, 26(2):* 195—224.

Haff, L. R. (1980) Empirical Bayes estimation of the multivariate normal covariance matrix, *The Annals of Statistics, 8:* 586—597.

Jagannathan, R. and Ma, T. (2003) Risk Reduction in Large Portfolios: Why imposing the wrong constraints helps. *The Journal of Finance, 58(4):*1651—1684

James, W., Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, 1:* 361—379.

Jobson, J.D., Korkie, B. (1980) Estimation of Markowitz efficient portfolios. *Journal of the American Statistical Association 75 (371):* 544—554.

Jobson, J.D., Korkie, B.M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance, 36:* 889—908.

Jobson, J.D., Korkie, B. (1989) A performance interpretation of multivariate tests of asset set intersection, spanning, and mean-variance efficiency. *Journal of Financial and Quantitative Analysis 24 (2):* 185—204.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis, 21:* 279—292.

Jorion, P. (1991). Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection, *Journal of Banking and Finance, 15:* 717—727

Kempf, A., Memmel, C. (2006). Estimating the global minimum variance portfolio. *Schmalenbach Business Review 58:* 332—248.

Ledoit, O. and Wolf, M. (2003,a). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, 10:* 603—621.

Ledoit, O. and Wolf, M. (2003,b). Honey, I shrunk the sample covariance matrix. *UPF Economics and Business Working Paper:* 691—702.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88(2):* 365—411.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance, 15:* 850—859.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics, 40(2):* 1024—1060.

Lettau, M. and Ludvigson, S. (2001) Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia are Time Varying. *Journal of Political Economy 109(6):* 1238—1287.

Levina, E., Rothman, A. and Zhu, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *The Annals of Applied Statistics (2):* 245—263.

Markowitz, H. (1952) Portfolio selection. *The Journal of Finance 7:* 77?91.

Markowitz, H. (1991) Foundations of portfolio theory. *The Journal of Finance 46:* 469—477.

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters, 1:*, 21—23.

Merton, R.C. (1980). On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics, 8:* 323—361.

Muirhead, R.J., (1987) Developments in eigenvalue estimation. *Advances in Multivariate Statistical* 277—288.

Okhrin, Y., Schmid, W. (2006). Distributional properties of portfolio weights. *Journal of Econometrics, 134:* 235—256.

Samuelson, P.A. (1970) The fundamental approximation theorem of portfolio analysis in terms of means, variances, and higher moments. *Review of Economical Studies 36:* 537—542.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science, 9(2)*: 277—293.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multi- variate normal distribution. *Proceedings of the Third Berkeley symposium on mathematical statistics and probability, 1:*197—206.

Sun, D., JO Berger, J.O. (1998) Reference priors with partial information. *Biometrika, 85(1):* 55—71.

Yang, R. and Berger, J.O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics 22:* 1195—1121.

Yu, B.W.T., Pang, W.K., Troutt, M.D., Hou, S.H. (2009) Objective comparisons of the optimal portfolios corresponding to different utility functions. *European Journal of Operational Research, 199:* 604—610.

List of 449 S&P500 stocks used

1 Agilent Technologies Inc (A)

2 Alcoa Inc (AA)

3 Apple Inc (AAPL)

4 AmerisourceBergen Corp (ABC)

5 Abbott Laboratories (ABT)

6 ACE Limited (ACE)

7 Actavis Inc (ACT)

8 Adobe Systems Inc (ADBE)

9 Analog Devices Inc (ADI)

10 Archer-Daniels-Midland Co (ADM)

11 Automatic Data Processing (ADP)

12 Alliance Data Systems Corp (ADS)

13 Autodesk Inc (ADSK)

14 Ameren Corp (AEE)

15 American Electric Power (AEP)

16 AES Corp (AES)

17 Aetna Inc (AET)

18 AFLAC Inc (AFL)

19 Allergan Inc (AGN)

20 American Intl Group Inc (AIG)

21 Apartment Investment & Mgmt (AIV)

22 Assurant Inc (AIZ)

23 Akamai Technologies Inc (AKAM)

24 Allstate Corp (ALL)

25 Altera Corp (ALTR)

26 Alexion Pharmaceuticals (ALXN)

27 Applied Materials Inc (AMAT)

28 AMETEK Inc (AME)

29 Amgen Inc (AMGN)

30 American Tower Corp A (AMT)

31 Amazon.com Inc (AMZN)

32 AutoNation Inc (AN)

33 Aon plc (AON)

34 Apache Corporation (APA)

35 Anadarko Petroleum Corp (APC)

36 Air Products & Chemicals Inc (APD)

37 Amphenol Corp A (APH)

38 Airgas Inc (ARG)

39 Allegheny Technologies Inc (ATI)

40 AvalonBay Communities Inc (AVB)

41 Avon Products (AVP)

42 Avery Dennison Corp (AVP)

43 American Express Co (AXP)

44 AutoZone Inc (AZO)

45 Boeing Company (BA)

46 Bank of America Corp (BAC)

47 Baxter International Inc (BAX)

48 Bed Bath & Beyond (BBBY)

49 BB&T Corporation (BBT)

50 Best Buy Co. Inc (BBY)

51 Bard (C.R.) Inc (BCR)

52 Becton Dickinson (BDX)

53 Franklin Resources (BEN)

54 Brown-Forman Corporation (BF-B)

55 Baker Hughes Inc (BHI)

56 BIOGEN IDEC Inc (BIIB)

57 The Bank of New York Mellon Corp (BK)

58 BlackRock (BLK)

59 Ball Corp (BLL)

60 Bemis Company (BMS)

61 Bristol-Myers Squibb (BMY)

62 Broadcom Corporation (BRCM)

63 Berkshire Hathaway (BRK-B)

64 Boston Scientific (BSX)

65 Peabody Energy (BTU)

66 BorgWarner (BWA)

67 Boston Properties (BXP)

68 Citigroup Inc (C)

69 CA Inc (CA)

70 ConAgra Foods Inc (CAG)

71 Cardinal Health Inc (CAH)

72 Cameron International Corp (CAM)

73 Caterpillar Inc (CAT)

74 Chubb Corp (CB)

75 CBRE Group (CBG)

76 Coca-Cola Enterprises (CCE)

77 Crown Castle International Corp (CCI)

78 Carnival Corp (CCL)

79 Celgene Corp (CELG)

80 Cerner (CERN)

81 Chesapeake Energy (CHK)

82 C. H. Robinson Worldwide (CHRW)

83 CIGNA Corp (CI)

84 Cincinnati Financial (CINF)

85 Colgate-Palmolive (CL)

86 Clorox Co (CLX)

87 Comerica Inc (CMA)

88 Comcast Corp (CMCSA)

89 CME Group Inc (CME)

90 Cummins Inc (CMI)

91 CMS Energy (CMS)

92 CenterPoint Energy (CNP)

93 CONSOL Energy Inc (CNX)

94 Capital One Financial (COF)

95 Cabot Oil & Gas (COG)

96 Coach Inc (COH)

97 Rockwell Collins (COL)

98 ConocoPhillips (COP)

99 Costco Co (COST)

100 Campbell Soup (CPB)

101 Salesforce.com (CRM)

102 Computer Sciences Corp (CSC)

103 Cisco Systems (CSCO)

104 CSX Corp (CSX)

105 Cintas Corporation (CTAS)

106 CenturyLink Inc (CTL)

107 Cognizant Technology Solutions (CTSH)

108 Citrix Systems (CTXS)

109 Cablevision Systems Corp (CVC)

110 CVS Caremark Corp (CVS)

111 Chevron Corp (CVX)

112 Dominion Resources (D)

113 Du Pont (EI) (DD)

114 Deere & Co (DE)

115 Quest Diagnostics (DGX)

116 D. R. Horton (DHI)

117 Danaher Corp (DHR)

118 Walt Disney Co (DIS)

119 Dollar Tree (DLTR)

120 Dun & Bradstreet (DNB)

121 Denbury Resources Inc (DNR)

122 Diamond Offshore Drilling (DO)

123 Dover Corp (DOV)

124 Dow Chemical (DOW)

125 Darden Restaurants (DRI)

126 DTE Energy Co (DTE)

127 DirecTV (DTV)

128 Duke Energy (DUK)

129 DaVita Inc (DVA)

130 Devon Energy Corp (DVN)

131 Electronic Arts (EA)

132 eBay Inc (EBAY)

133 Ecolab Inc (ECL)

134 Consolidated Edison (ED)

135 Equifax Inc (EFX)

136 Edison Int'l (EIX)

137 Estee Lauder Cos (EL)

138 EMC Corp (EMC)

139 Eastman Chemical (EMN)

140 Emerson Electric (EMR)

141 EOG Resources (EOG)

142 Equity Residential (EQR)

143 EQT Corporation (EQT)

144 Express Scripts (ESRX)

145 Essex Property Trust (ESS)

146 Ensco plc (ESV)

147 E-Trade (ETFC)

148 Eaton Corp (ETN)

149 Entergy Corp (ETR)

150 Edwards Lifesciences (EW)

151 Exelon Corp (EXC)

152 Expeditors Int'l (EXPD)

153 Ford Motor (F)

154 Fastenal Co (FAST)

155 Freeport-McMoran Cp & Gld (FCX)

156 Family Dollar Stores (FDO)

157 FedEx Corporation (FDX)

158 FirstEnergy Corp (FE)

159 F5 Networks (FFIV)

160 Fidelity National Information Services
(FIS)

161 Fiserv Inc (FISV)

162 Fifth Third Bancorp (FITB)

163 FLIR Systems (FLIR)

164 Fluor Corp (FLR)

165 Flowserve Corporation (FLS)

166 FMC Corporation (FMC)

167 Fossil Inc (FOSL)

168 Twenty-First Century Fox Inc (FOXA)

169 Forest Laboratories (FRX)

170 FMC Technologies Inc (FTI)

171 Frontier Communications (FTR)

172 AGL Resources Inc (GAS)

173 Gannett Co (GCI)

174 General Dynamics (GD)

175 General Electric (GE)

176 General Growth Properties Inc (GGP)

177 Graham Holdings Co (GHC)

178 Gilead Sciences (GILD)

179 General Mills (GIS)

180 Corning Inc (GLW)

181 Keurig Green Mountain Inc (GMCR)

182 GameStop Corp (GME)

183 Genworth Financial Inc (GNW)

184 Genuine Parts (GPC)

185 Gap (The) (GPS)

186 Garmin Ltd (GRMN)

187 Goldman Sachs Group (GS)

188 Goodyear Tire & Rubber (GT)

189 Grainger (W.W.) Inc (GWW)

190 Halliburton Co (HAL)

191 Harman Int'l Industries (HAR)

192 Hasbro Inc (HAS)

193 Huntington Bancshares (HBAN)

194 Hudson City Bancorp (HCBK)

195 REIT (HCN)

196 HCP Inc (HCP)

197 Home Depot (HD)

198 Hess Corporation (HES)

199 Hartford Financial Svc. Gp (HIG)

200 Harley-Davidson (HOG)

201 Honeywell Int'l Inc (HON)

202 Starwood Hotels & Resorts (HOT)

203 Helmerich & Payne (HP)

204 Hewlett-Packard (HPQ)

205 Block H&R (HRB)

206 Hormel Foods Corp (HRL)

207 Harris Corporation (HRS)

208 Hospira Inc (HSP)

209 Host Hotels & Resorts (HST)

210 The Hershey Company (HSY)

211 Humana Inc (HUM)

212 International Bus. Machines (IBM)

213 International Flav/Frag (IFF)

214 Intel Corp (INTC)

215 Intuit Inc (INTU)

216 International Paper (IP)

217 Interpublic Group (IPG)

218 Ingersoll-Rand PLC (IR)

219 Iron Mountain Inc (IRM)

220 Intuitive Surgical Inc (ISRG)

221 Illinois Tool Works (ITW)

222 Invesco Ltd (IVZ)

223 Jabil Circuit (JBL)

224 Johnson Controls (JCI)

225 Jacobs Engineering Group (JEC)

226 Johnson & Johnson (JNJ)

227 Juniper Networks (JNPR)

228 Joy Global Inc (JOY)

229 JPMorgan Chase & Co (JPM)

230 Nordstrom (JWN)

231 Kellogg Co (K)

232 KeyCorp (KEY)

233 Kimco Realty (KIM)

234 KLA-Tencor Corp (KLAC)

235 Kimberly-Clark (KMB)

236 Carmax Inc (KMX)

237 Coca Cola Co (KO)

238 Kroger Co (KR)

239 Kohl's Corp (KSS)

240 Kansas City Southern Inc (KSU)

241 Loews Corp (L)

242 L Brands Inc (LB)

243 Leggett & Platt (LEG)

244 Lennar Corp (LEN)

245 Laboratory Corp of America Holding (LH)

246 L-3 Communications Holdings (LLL)

247 Linear Technology Corp (LLTC)

248 Lilly (Eli) & Co (LLY)

249 Legg Mason (LM)

250 Lockheed Martin Corp (LMT)

251 Lincoln National (LNC)

252 Lowe's Cos (LOW)

253 Lam Research (LRCX)

254 Leucadia National Corp (LUK)

255 Southwest Airlines (LUV)

256 Macy's Inc (M)

257 Macerich Co (MAC)

258 Marriott Int'l (MAR)

259 Masco Corp (MAS)

260 Mattel Inc (MAT)

261 McDonald's Corp (MCD)

262 Microchip Technology (MCHP)

263 McKesson Corp (MCK)

264 Moody's Corp (MCO)

265 Mondelez International Inc (MDLZ)

266 Medtronic Inc (MDT)

267 MetLife Inc (MET)

268 McGraw Hill Financial Inc (MHFI)

269 Mohawk Industries Inc (MHK)

270 McCormick & Co (MKC)

271 Marsh & McLennan (MMC)

272 3M Co (MMM)

273 Monster Beverage (MNST)

274 Altria Group Inc (MO)

275 Monsanto Co (MON)

276 The Mosaic Company (MOS)

277 Merck & Co (MRK)

278 Marathon Oil Corp (MRO)

279 Morgan Stanley (MS)

280 Microsoft Corp (MSFT)

281 Motorola Solutions Inc (MSI)

282 M&T Bank Corp (MTB)

283 Micron Technology (MU)

284 Murphy Oil (MUR)

285 MeadWestvaco Corporation (MWV)

286 Mylan Inc (MYL)

287 Noble Energy Inc (NBL)

288 Nabors Industries Ltd (NBR)

289 NASDAQ OMX Group (NDAQ)

290 Noble Corp (NE)

291 NextEra Energy Resources (NEE)

292 Newmont Mining Corp (NEM)

293 NetFlix Inc (NFLX)

294 Newfield Exploration Co (NFX)

295 NiSource Inc (NI)

296 NIKE Inc (NKE)

297 Northrop Grumman Corp (NOC)

298 National Oilwell Varco Inc (NOV)

299 NRG Energy (NRG)

300 Norfolk Southern Corp (NSC)

301 NetApp (NTAP)

302 Northern Trust Corp (NTRS)

303 Northeast Utilities (NU)

304 Nucor Corp (NUE)

305 Nvidia Corporation (NVDA)

306 Newell Rubbermaid Co (NWL)

307 Owens-Illinois Inc (OI)

308 ONEOK (OKE)

309 Omnicom Group (OMC)

310 Oracle Corp (ORCL)

311 O'Reilly Automotive (ORLY)

312 Occidental Petroleum (OXY)

313 Paychex Inc (PAYX)

314 People's United Bank (PBCT)

315 Pitney-Bowes (PBI)

316 PACCAR Inc (PCAR)

317 PG&E Corp (PCG)

318 Plum Creek Timber Co (PCL)

319 Priceline.com Inc (PCLN)

320 Precision Castparts (PCP)

321 Patterson Companies (PDCO)

322 Public Serv. Enterprise Inc (PEG)

323 PepsiCo Inc (PEP)

324 PETsMART Inc (PETM)

325 Pfizer Inc (PFE)

326 Principal Financial Group (PFG)

327 Procter & Gamble (PG)

328 Progressive Corp (PGR)

329 Parker-Hannifin (PH)

330 Pulte Homes Inc (PHM)

331 PerkinElmer (PKI)

332 ProLogis (PLD)

333 Pall Corp (PLL)

334 PNC Financial Services (PNC)

335 Pentair Ltd (PNR)

336 Pinnacle West Capital (PNW)

337 Pepco Holdings Inc (POM)

338 PPG Industries (PPG)

339 PPL Corp (PPL)

340 Perrigo (PRGO)

341 Prudential Financial (PRU)

342 Public Storage (PSA)

343 PVH Corp (PVH)

344 Quanta Services Inc (PWR)

345 Praxair Inc (PX)

346 Pioneer Natural Resources (PXD)

347 QUALCOMM Inc (QCOM)

348 Ryder System (R)

349 Reynolds American Inc (RAI)

350 RDCRowan Cos (RDC)

351 Regeneron Pharmaceuticals Inc (REGN)

352 Regions Financial Corp (RF)

353 Robert Half International (RHI)

354 Red Hat Inc (RHT)

355 Transocean Ltd (RIG)

356 Polo Ralph Lauren Corp (RL)

357 Rockwell Automation Inc (ROK)

358 Roper Industries (ROP)

359 Ross Stores Inc (ROST)

360 Range Resources Corp (RRC)

361 Republic Services Inc (RSG)

362 Raytheon Co (RTN)

363 Starbucks Corp (SBUX)

364 SCANA Corp (SCG)

365 Charles Schwab (SCHW)

366 Sealed Air Corp(New) (SEE)

367 Sherwin-Williams (SHW)

368 Sigma-Aldrich (SIAL)

369 Smucker (J.M.) (SJM)

370 Schlumberger Ltd (SLB)

371 Snap-On Inc (SNA)

372 SanDisk Corporation (SNDK)

373 Southern Co (SO)

374 Simon Property Group Inc (SPG)

375 Staples Inc (SPLS)

376 Stericycle Inc (SRCL)

377 Sempra Energy (SRE)

378 SunTrust Banks (STI)

379 St Jude Medical (STJ)

380 State Street Corp (STT)

381 Seagate Technology (STX)

382 Constellation Brands (STZ)

383 Stanley Black & Decker (SWK)

384 Southwestern Energy (SWN)

385 Safeway Inc (SWY)

386 Stryker Corp (SYK)

387 Symantec Corp (SYMC)

388 Sysco Corp (SYY)

389 AT&T Inc (T)

390 Molson Coors Brewing Company (TAP)

391 TECO Energy (TE)

392 Integrys Energy Group Inc (TEG)

393 Target Corp (TGT)

394 Tenet Healthcare Corp (THC)

395 Tiffany & Co (TIF)

396 TJX Companies Inc (TJX)

397 Torchmark Corp (TMK)

398 Thermo Fisher Scientific (TMO)

399 T. Rowe Price Group (TROW)

400 The Travelers Companies Inc (TRV)

401 TSCOTractor Supply Co (TSCO)

402 Tyson Foods (TSN)

403 Tesoro Petroleum Co (TSO)

404 Total System Services (TSS)

405 Time Warner Inc (TWX)

406 Texas Instruments (TXN)

407 Textron Inc (TXT)

408 Tyco International (TYC)

409 United Health Group Inc (UNH)

410 Unum Group (UNM)

411 Union Pacific (UNP)

412 United Parcel Service (UPS)

413 Urban Outfitters (URBN)

414 U.S. Bancorp (USB)

415 United Technologies (UTX)

416 Varian Medical Systems (VAR)

417 V.F. Corp (VFC)

418 Valero Energy (VLO)

419 Vulcan Materials (VMC)

420 Vornado Realty Trust (VNO)

421 Verisign Inc (VRSN)

422 Vertex Pharmaceuticals Inc (VRTX)

423 Ventas Inc (VTR)

424 Verizon Communications (VZ)

425 Walgreen Co (WAG)

426 Waters Corporation (WAT)

427 Western Digital (WDC)

428 Wisconsin Energy Corporation (WEC)

429 Wells Fargo (WFC)

430 Whole Foods Market (WFM)

431 Whirlpool Corp (WHR)

432 WellPoint Inc (WLP)

433 Waste Management Inc (WM)

434 Williams Cos (WMB)

435 Wal-Mart Stores (WMT)

436 Weyerhaeuser Corp (WY)

437 Wynn Resorts Ltd (WYNN)

438 United States Steel Corp (X)

439 Cimarex Energy Co (XEC)

440 Xcel Energy Inc (XEL)

441 XL Capital (XL)

442 Xilinx Inc (XLNX)

443 Exxon Mobil Corp (XOM)

444 Dentsply International (XRAY)

445 Xerox Corp (XRX)

446 Yahoo Inc (YHOO)

447 Yum! Brands Inc (YUM)

448 Zions Bancorp (ZION)

449 Zimmer Holdings (ZMH)

Declaration of Authorship

I hereby confirm that I have authored this Master's thesis independently and without use of others than the indicated sources. All passages which are literally or in general matter taken out of publications or other sources are marked as such.

Berlin, April 20, 2015

Svetlana Bykovskaya